

Technical Report 15-002

Potential of Using Mobile Phone Data to Assist in Mission Analysis and Area of Operations Planning

Andrew C. Lee

U.S. Military Academy, West Point NY

August 2015

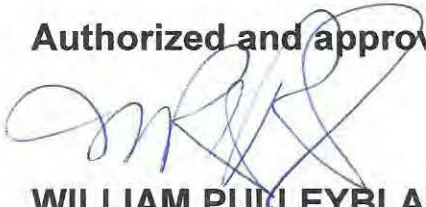


**United States Military Academy
Network Science Center**

Approved for public release; distribution is unlimited.

**U.S. Military Academy
Network Science Center**

Authorized and approved for distribution:



**WILLIAM PULLEYBLANK, Ph.D.
Network Science Center Chair**



**LTC ANTHONY JOHNSON, Ph.D.
Director, Network Science Center**

Technical review by

Nakeya Williams, Ph.D., Department of Mathematical Sciences, U.S. Military Academy

COL Mason Crow, Department of Mathematical Sciences, U.S. Military Academy

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by the U.S. Military Academy Network Science Center. Please address correspondence concerning distribution of reports to: Network Science Center, U.S. Military Academy, 646 Swift Road, West Point, NY 10996

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Military Academy Network Science Center.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 12 July 2015		2. REPORT TYPE Technical Report		3. DATES COVERED (From - To) April 2015-June 2015	
4. TITLE AND SUBTITLE Potential of Using Mobile Phone Data to Assist in Mission Analysis and Area of Operations Planning				5a. CONTRACT NUMBER n/a	
				5b. GRANT NUMBER n/a	
				5c. PROGRAM ELEMENT NUMBER n/a	
6. AUTHOR(S) MAJ Andrew C. Lee				5d. PROJECT NUMBER	
				5e. TASK NUMBER n/a	
				5f. WORK UNIT NUMBER n/a	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Mathematical Sciences, U.S. Military Academy				8. PERFORMING ORGANIZATION REPORT NUMBER n/a	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) USMA Network Science Center				10. SPONSOR/MONITOR'S ACRONYM(S) USMA NSC	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) 15-002	
12. DISTRIBUTION / AVAILABILITY STATEMENT Unlimited Distribution					
13. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.					
14. ABSTRACT The way a unit defines an area of operation (AO) can restrict or facilitate a unit's ability to influence the population; however, units may not have access to accurate census data. An AO developed through a cursory mission analysis could misrepresent current tribal boundaries, population densities, district boundaries, and residential areas. Recent research has demonstrated that given anonymized call data records of mobile subscribers, units may be able to circumvent the need for an accurate census to feed into the mission analysis in determining an optimal AO. Furthermore, mobile data also provides the added benefit of allowing us to assess changing population patterns over time.					
15. SUBJECT TERMS Mobile phone data, call data records, area of operation, census					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UL	18. NUMBER OF PAGES 10	19a. NAME OF RESPONSIBLE PERSON Tish Torgerson
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED			19b. TELEPHONE NUMBER (include area code) 845-938-8238

Potential of Using Mobile Phone Data to Assist in Mission Analysis and Area of Operations Planning

Andrew C. Lee

1 Introduction

An Area of Operations (AO) is a contiguous or noncontiguous operational area activated by the commander at a set date and time based on mission and situation considerations. AOs are typically fixed once a unit arrives in theater and therefore, its boundaries can play a significant role in overall mission success. According to Joint Publication 3-0, "Areas of Operations do not typically encompass the entire operational area of the joint force commander, but should be large enough for component commanders to accomplish their missions and protect their forces [10]. According to doctrine, the Joint Force Land Component Commander (JFLCC) assigns their subordinates their AOs and these units in turn assigns their subordinates' AOs down to the company level based on several factors like Mission, Enemy, Terrain and Weather, Troops and Support Available, Time Available, and Civil Considerations (METT-TC) [6]. There are other common considerations for determining an AO, to include "Command and Control (C2), the information environment, intelligence requirements, communications support, protection, security, lines of communications, terrain management, movement control, airspace control, surveillance, reconnaissance, air and missile defense, personnel recovery, targeting and fires, inter-organizational interactions, and environmental issues" [10].

FM 3-90 outlines the following potential restrictions from an AO:

- 1) Restricts outside units from moving through the AO
- 2) Restricts outside units from use of fires

In order to get past these restrictions, the owning unit in the AO must provide clearance [6]. Yet this coordination between units can result in both loss of time and potential administrative or technical hurdles. Furthermore, although some units attempt to relax these restrictions with explicit prior coordination, unit commanders typically tend to revert back to focusing only on their respective AOs. As a result, AO boundaries can significantly impact or hamper the ability of the commander to shape the environment. For example, if a critical tribe is located in an adjacent unit's AO, this could limit the commander's ability to engage in sphere of influence (SOI) discussions. As another example, a High Value Target (HVT) that might reside in an adjacent unit's AO may not receive the level of attention or priority desired. A commander can also request to change his or her AO, but in my experience this is rarely done.

Since an AO can both restrict and facilitate a unit's ability to influence the population, how a unit carves out an AO can ultimately achieve mission success. Yet, units usually rely on outdated data in deployed environments. For example, the last population census for Afghanistan was completed in 1979 [16] so the AO developed through a cursory analysis of civil considerations could misrepresent current tribal boundaries, population densities, district boundaries, and residential areas. Reliable and accurate mapping of populations require robust census surveys which can take years and significant resources to accomplish. Additionally, spatial changes due to disasters, epidemics, or violence can alter the population map in a short amount of time. With a high number of users in most countries, mobile

phones offer a promising alternative in gathering some of the census data required without the logistical, time, and financial burden.

Given the call data records of mobile subscribers, one would be able to circumvent the need for an accurate census to feed into the METT-TC analysis in determining an optimal AO. Furthermore, mobile data also provides the added benefit of allowing us to assess changing population patterns over time. This paper aims to provide some background on mobile phone data and how it might be used to assist commanders in carving out a proper AO in countries that lack sufficient census data.

1.1 Motivation

Due to lack of personnel and resources, most units do not fully complete the scrupulous METT-TC analysis required for an accurate depiction of the environment prior to deployment. As stated before, the last population census in Afghanistan was completed in 1979. With such outdated and unreliable data, estimates of at risk populations or ethnic and tribal boundaries could lead to wrong conclusions for disaster assessments, intervention planning, and conflict relief planning. More importantly, even if there was an accurate census, migration and displacement of populations from war could significantly alter the demographic and economic makeup of the AO in a relatively short amount of time. As seen in Iraq, mass executions and targeting of Christian populations have forced a majority of them to leave the country.

Questions with respect to AO analysis include: 1) Are there other sources of information that could provide more accurate population and demographic data for a deploying unit entering a country for the first time? 2) Is there a way to update this data at low cost given significant changes to the population due to displacements and migrations? Mobile phone data can help overcome the limitations of a census especially in hostile or restricted areas. Census data usually includes information on population, gender, age, race or tribe, number of homes, family size, employment status, language spoken, and education attainment [15]. Of these, mobile phone data has the potential to provide good estimates for population density, tribal areas, and language. Furthermore, research has shown that analysis of cell phone data usage patterns can allow inferences to be made about socioeconomic welfare, mobility patterns, and financial activity [2].

Since data from Afghanistan telecom companies is currently difficult to obtain, we can use data on the country of Senegal through the Data for Development's innovation challenge made available by Sonatel and the Orange Group as an interim focus for research and as a proof of concept.

2 Research Goals

Our research goal is to determine if mobile phone data can address the following questions in a developing nation:

- 1) Approximately how many people are there in each region of the country?
- 2) What do the tribal or ethnic boundaries look like and how do these compare with current maps?
- 3) Are there underlying homophily preferences that can be used to predict mobility?
- 4) How does the population distribution change over time in relation to the week, season, or particular events like violent attacks?

By addressing these questions with mobile phone data, a unit could circumvent the census approach and still gain an accurate depiction of the population to determine the most favorable AO boundaries.

3 Literature Review

3.1 Mapping Populations Using Mobile Phone Data

The first requirement in determining an AO is to determine population density in order to ensure the unit assigned to the area has the proper force ratio to deal with the population. One approach used by Deville et al. to address the question of population density used 1 billion mobile phone call records from Portugal and France to compare the spatial estimations of population densities at national scales with estimates from more traditional population mapping methods [5]. The results of their population density datasets is shown in figure 1.

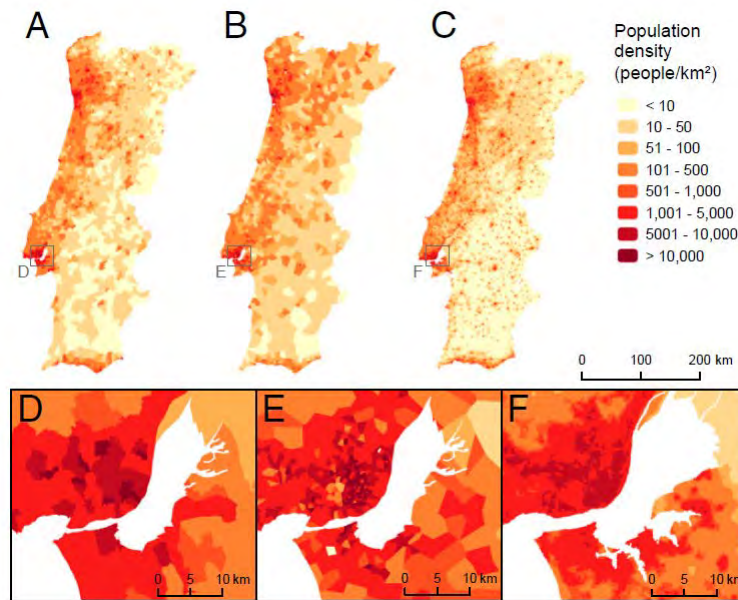


Figure 1. Comparison of predicted population density datasets with baseline data for mainland Portugal. A) Population density as calculated from a national census, B) Population density at the level of Voronoi polygons using the MP method, and C) Population density at the level of 100 x 100 m grid squares as estimated by the RS method. D-F) Close ups around Lisbon [5]

At the national scale, both mobile phone (MP) data based approaches and the Remote Sensing (RS) method they used closely resembled the baseline census population data. At the microscopic city level however, more contrasts exist mainly due to the reliance of the MP data on the spatial density of cell towers. In urban areas, the density of towers is typically higher which can provide more accuracy. The authors also demonstrated that mobile data allows us to assess changing population distributions over short time intervals, paving the way for near real time understanding of patterns. For example, they showed seasonal changes in population distribution and in particular, how people from densely populated areas tend to travel to low density and recreational locations during holidays or weekends as seen in Figure 2.

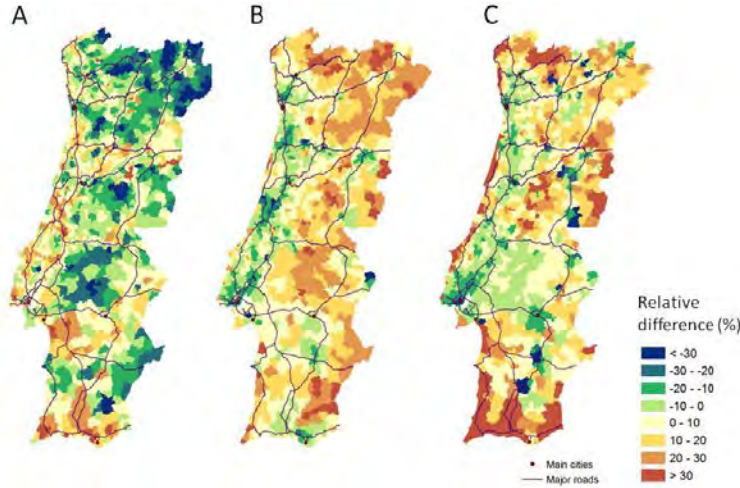


Figure 2. Relative difference in predicted population density for different time periods in Portugal. A) Difference between day and night with brown colors indicating a higher population density during the day, B) difference between weekend and weekdays with brown colors indicating a higher population density during weekends, C) difference between the main holiday period (July and August) and the working period (Nov-May) with brown colors indicating a higher population density during the holidays [5]

3.2 Uncovering Human Mobility Patterns

Another advantage mobile phone data could provide for AO analysis would be in determining predictable movement patterns. Gonzalez et al. studied individual human mobility patterns using anonymized mobile data. They studied the trajectory of 100,000 users over a six-month period and discovered that humans show a high degree of temporal and spatial regularity with a significant probability of returning to a few set of locations [8]. In other words, humans tend to be predictable in their travel patterns that can be represented in a single spatial probability distribution. By measuring the distance between the user's positions at consecutive calls, they found that the distribution of displacements over all users is described by a truncated power law:

$$P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} e^{(-\Delta r/\kappa)} \quad \text{Equation 1}$$

Using data from Boston and Los Angeles, research by Desu et al. focused on addressing the question: Do people from segregated areas tend to stay in these areas? She showed through visual representations that the answer is indeed yes; there are obvious homophily preferences inherent with different races and there are only a few regions in a city where there is a true mix of different cultures and ethnicities [4].

Applying this approach to the country of Afghanistan would be tremendously beneficial especially since a sizeable portion of the population are nomads, changing location based on season. A proper AO assessment would need to take this into consideration based on the seasonal movement patterns. In addition, the different ethnicities in Afghanistan do indeed show homophily preferences.

3.3 Predicting Gender Based on CDR

Finally, mobile phone data could also reveal age and gender estimates that might be useful in mission analysis, especially if trying to determine an estimate for the number of military age males in an area. There were several studies that attempted to infer gender or age from calling activity. In a study on landlines, Smoreda et al. found significant differences between genders and found that women tend to call more often than men and more interestingly, the gender of the callee can help predict the duration of the call [13]. Building on this research on landline calls, Frias-Martinez et al. extended this research to

mobile data and proposed a way to infer the gender based on different behavioral variables. Their research showed that they could achieve a success rate of prediction between 70% and 80% [7]. Furthermore, Jahani et al. attempted to predict gender based on mobile data using Bandicoot behavioral indicators. As motivation for their research, they mentioned that countries like Afghanistan are still using census data from 1973 and no real accurate data currently exists. Jahani et al. showed that gender prediction can work to an extent but future work on further machine learning applications should be used to compare across countries in order to predict more meaningful socioeconomic information [9].

4 Data

At the end of September 2014, there were 23,114,471 Global System for Mobile Communications (GSM) subscribers with a population coverage of over 89% according to the Ministry of Communication and Information Technology in Afghanistan [11]. The growth of total mobile phone users in Afghanistan since 2002 was impressive given that no such infrastructure existed beforehand, with the biggest increase from 2009 to 2010 of 70% [1]. Despite this promising trend for mobile phone data, there are also unique considerations in Afghanistan where the Taliban have decreed in the past that cell towers be turned off nightly or be destroyed for two main reasons: 1) Due to fear of tracking and monitoring by Coalition forces, and 2) to limit local tips to Coalition forces [14]. This could lead to disproportional penetration in certain areas of the population like the Helmand province in the South. Furthermore, since night time data is typically used to analyze residential locations and density, night time blackouts could present a challenge for this type of analysis.

Ideally, mobile data from Afghanistan's cell phone carriers would provide the best data; however, in the absence of such data from leading telecom companies Etisalat, Roshan, MTN, and Afghan Wireless, a recommendation would be to use comparable data from another country in order to demonstrate proof of concept.

The Orange Corporation's Data for Development is an innovation challenge on Big Data for the purposes of societal development (see <http://www.d4d.orange.com/en/home>). Anonymous data, extracted from the mobile network is made available to international research laboratories to contribute to the health, agriculture, transportation/urban planning, energy, and national statistics of the country being studied. To be eligible for the project, the participant has to be an academic university, faculty, non-profit entity, or scientific research institution, and the participant must agree to keep the data confidential perpetually. This past year, the Data for Development's innovation challenge made available anonymized Call Detail Records (CDR) from the country of Senegal for a short period of time. The datasets were based on phone calls and text exchanges between more than 9 million of Sonatel's customers in Senegal collected from January 1 to December 31, 2013. Being able to map population, mobility, and behavioral indicators from this data would not only demonstrate proof of concept but would also serve as motivation for future research in applying this to AO mission analysis. More details about the data provided during the NetMob Conference is available in Appendix A.

5 Challenges and Risks

Depending on the country being studied, telecom companies may be hesitant to provide data even if it is anonymized. Privacy concerns discourage the use of mobile phone data as it relates to issues of personal freedom and ethics. However, by using anonymized data and call activity aggregated by towers, the

privacy of the mobile phone users can be guaranteed. Research done in this field has shown that mobile phone companies, provided the right incentives, can share the data in this anonymized and aggregated format [2].

Another challenge is the difference across countries with varying cultures and behaviors. For example, in Afghanistan, one might need to make some adjustments to predictions based on mobile phone user density, phone use behaviors, and unequal penetration rates. Furthermore, a big potential issue with data from third world nations is the rampant use of prepaid cards. Despite these challenges, the rapidly growing use of mobile phones and ongoing research should help overcome some of these challenges.

6 Future Work

The use of mobile phone data to analyze a population in a developing country like Afghanistan could make an immediate impact in the quality of AO mission analysis. In addition, another application could be in transportation infrastructure development based on mobility patterns derived from mobile phone data. Another related research question to this paper would be: What is the ideal area of operations in terms of ethnicity coverage? For example, in a country with distinct and clearly defined ethnic boundaries, would it be better if a commander owned an AO with portions of each ethnic neighborhood in order to promote diversity and unification, or would it be better to own homogeneous AOs with respect to ethnic regions?

As one of the next frontiers of big data applications, other telecom companies are sure to follow suit and offer anonymized data in the form of competitions [2]. Although the timeline for the next Orange Telecom project is not yet known, other opportunities currently exist. For example, the Data for Climate Action project sponsored by the United Nations “aims to spur greater action on climate change by accelerating innovations in data-driven climate solutions and helping to build resilience around the world.” [15]. This project will also include a research challenge in early 2016.

These research competitions can also provide an excellent opportunity for the United States Military Academy (USMA)’s Network Science Center (NSC) to work on real world big data applications. The NSC was founded at West Point to bring together “service members, civilians, and cadets to research and develop significant contributions in the study of network representations of physical, biological, and social phenomena leading to models” [18]. For the Cadets, this research could also contribute to a thesis for Mathematical Sciences Majors and Operations Research Majors who take either a semester long or year-long thesis course as a requirement for their undergraduate degree. From the NetMob conference, I learned that with sufficient technical expertise, it is possible in 2 days to download the data and to run a quick analysis on the mobile data to analyze topics like mobility, migration patterns, socioeconomic status, and how to predict major events based on cell phone activity.

Appendix A

NetMob Conference Summary:

In order to learn more about mobile phone data analysis, I participated in the NetMob conference held at MIT's Media Lab from April 7-9, 2015. This was a conference on the scientific analysis of mobile phone datasets which also involved a school and hackathon this year with access to data from the country of Senegal. This was made possible through the Data for Development's innovation challenge made available by Sonatel and the Orange Group. In 2011, the NetMob conference was focused on the country of Cote D'Ivoire. On April 7th 2015, I participated in the Hackathon program and from the 8-9th of April, participated in the scientific conference.

The Hackathon short course allowed us to work with the actual data. We signed a confidentiality agreement that limited us to having access to the data for only the duration of the Hackathon. The datasets were based on Call Detail Records (CDR) of phone calls and text exchanges between more than 9 million of Sonatel's customers in Senegal collected from January 1 to December 31, 2013.

Three separate data sets were constructed and made available to the participants:

1) Antenna to Antenna traffic for 1,666 antenna towers on an hourly basis. Monthly voice traffic between antenna towers were structured as follows:

- a. Timestamp: day and hour
- b. Outgoing_site_id: ID of site the call originated from
- c. Incoming_site_id: ID of site receiving the call
- d. Number_of_calls: The total number of calls between these two sites during this hour.
- e. Total_Call_Duration: The total duration of all calls between these two sites during this hour in seconds.

For text traffic, the data was structured slightly differently:

- a. Timestamp: day and hour
- b. Outgoing_site_id: ID of site the text originated from
- c. Incoming_site_id: ID of site receiving the text
- d. Number_of_sms: The total number of texts between these two sites during this hour.

2) Fine grained mobility data on a rolling 2 week basis with about 300,000 randomly sampled users.

For 25 two week periods for 300,000 randomly selected users, this data provided:

- a. User_id: Selected User's random ID
- b. Timestamp: 24 h format YYYY-MM-DD-HH:M0:00 (the second digits of the minutes and all the seconds of the timestamps were replaced with zeros for confidentiality).
- c. Site_id: ID of the antenna site.

3) One year coarse grained mobility data at 123 arrondissements (administrative districts) for about 150,000 randomly sampled users.

Only users having more than 75% days with interactions yearly were selected. This data provided:

- a. User_id: Selected User's random ID
- b. Timestamp: 24 h format YYYY-MM-DD-HH:M0:00 (the second digits of the minutes and all the seconds of the timestamps were replaced with zeros for confidentiality).
- c. Arrondissement_id: Arrondissement code.

Datasets 2 and 3 also came with bandicoot behavioral indicators at the individual level computed from CDRs using the bandicoot toolbox [3]. Behavioral indicators include the mean and standard error of the following: user's active days, duration of calls, entropy of contacts for calls, entropy of contacts for texts, entropy of places for call and text, interactions per contact for call and text, and the inter-events between calls and texts.

The course also provided an introduction to data science tools. Luc Rocher from MIT showed us how to navigate on the NetMob virtual machine implementing telnet and putty.org for those using a PC. He showed us how to navigate on the server using UNIX. Functions covered were how to move, copy, delete, list the data, see the filesize, and how to unzip files.

We were also given access to a file that gave the latitude and longitude of antenna sites with some random deviation built in along with the arrondissement codes as shown in Figure 3 below. Additionally, the arrondissement shapefile provided the geometrical center of the arrondissement coordinates, its code, and the names of the region, department, municipality/city, and arrondissement.

Due to confidentiality agreements, the data could not be taken outside of the workshop.



Figure 3. Antenna Towers Coordinates in Senegal

REFERENCES

- [1] Baharustani, R. (May 2013). Study of Afghan Telecom. Research, Planning & Policy Directorate, AISA.
- [2] Blondel, V. D., Decuyper, A., and Krings, G. (11 February 2015). A survey of results on mobile phone datasets analysis.
- [3] De Montjoye, Y-A., Rocher, L., and Pentland, A. (2014). Bandicoot – a python toolbox to learn from metadata. Retrieved May 15, 2015 from <http://bandicoot.mit.edu/docs>.
- [4] Desu, S., Alexander, L., and Gonzalez, M. (April 2015). Untangling the Effects of Residential Segregation on Individual Mobility. Paper presented at the NetMob Conference in Cambridge, MA.
- [5] Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., Tatem, A.J.: Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* 111(45), 15888–15893 (2014)
- [6] Field Manual 3-90. (July 2001). Tactics. Washington, D.C.: Headquarters, Department of the Army, 2-4.
- [7] Frias-Martinez, V., Frias-Martinez, E., Oliver, N.: A gender-centric analysis of calling behavior in a developing economy using call detail records. In: *AAAI Spring Symposium: Artificial Intelligence for Development* (2010)
- [8] Gonzalez, M. C., Hidalgo, C. A., Barabasi, A-L.: Understanding Individual Human Mobility Patterns. *Nature* 453, 779-782 (2008).
- [9] Jahani, E., Sundsoy, P. R., Bjelland, J., Iqbal, Asif., Pentland, A., and Montjoye, Y. (April 2015). Predicting Gender from Mobile Phone Metadata. Paper presented at the NetMob Conference in Cambridge, MA.
- [10] Joint Publication 3.0. (August 2011). Joint Operations. Washington, D.C.: Headquarters, Department of the Army.
- [11] Ministry of Communications and Information Technology, Islamic Republic of Afghanistan (2006). Statistics. Retrieved April 21, 2015 from <http://www.mcit.gov.af/en>.
- [12] Simini, F., Gonzalez, M.C., Maritan, A., Barabasi, A.-L.: A universal model for mobility and migration patterns. *Nature* 484(7392), 96–100 (2012)
- [13] Smoreda, Z., Licoppe, C.: Gender-specific use of the domestic telephone. *Social Psychology Quarterly* 63(3), 238–252 (2000) [12]

- [14] Trofimov, Y. (22 March 2010). Cell Towers Bow to Taliban Threat. Retrieved April 23, 2015 from <http://www.wsj.com>.
- [15] United Nations. Data for Climate Action. Retrieved June 24, 2015 from <http://www.dataforclimateaction.org/the-challenge>.
- [16] U.S. Census Bureau. (2013). *American Fact Finder*. Retrieved June 21, 2015 from <http://www.factfinder.census.gov/>.
- [17] U.S. Census Bureau. (2013). *International Programs: Afghanistan*. Retrieved April 17, 2015 from <http://www.census.gov/>.
- [18] United States Military Academy. Network Science Center. Retrieved June 24, 2015 from <http://www.westpoint.edu/nsc/SitePages/About.aspx>.